NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Czech Academic and Research
Discovery Services
CARDS

About the cost-benefit analysis

Hana Heringová

Webinar „Results of a cost-benefit analysis of the implementation of PIDs in the Czech Republic“

The EOSC CZ Training Centre

April 15, 2025

Co-funded by
the European Union

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

Why a cost-benefit analysis

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- NTK has received funding from the CARDS* project for **2023 – 2028** to create and fund:

National Centre for Persistent Identifiers

- Centre's Mission:

Provide methodological, administrative and financial support of adoption of PIDs compliant with international standards (ORCID iD, DOI, ISSN and others) in the Czech R&D&I

*CARDS (Czech Academic and Research Discovery Services) - Individual system project no CZ.02.01.01/00/22_004/0004342, Operational Programme Jan Amos Komenský

Why a cost-benefit analysis

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- **National Centre for Persistent Identifiers** runs
  - **National ISSN Centre**
  - **National DOI Centre** / national DataCite consortium
  - **National ORCID Centre** / national ORCID consortium
- Both consortia intended for up to 110 institutions by 2028;
  NTK is able to cover <u>all membership fees</u> until the end of CARDS project

NEED: assess and measure the impact of the funding

NTK comissioned a cost-benefit analysis of PID implementation
with MoreBrains Cooperative

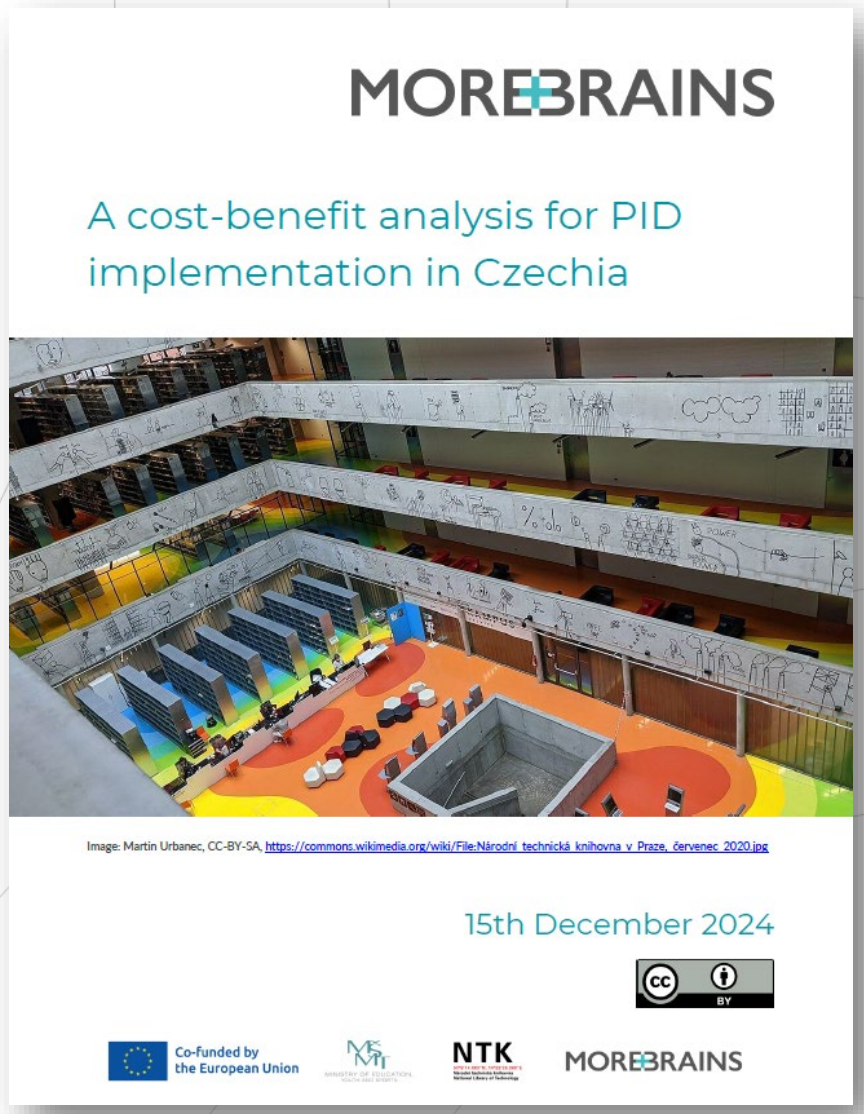
About the cost-benefit analysis

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Conducted in 2H 2024,
next phase 2028

English version:
https://doi.org/10.48813/x4vd-yj13

Czech version:
https://doi.org/10.48813/zm93-wy56



Reach out

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

identifikátory.cz

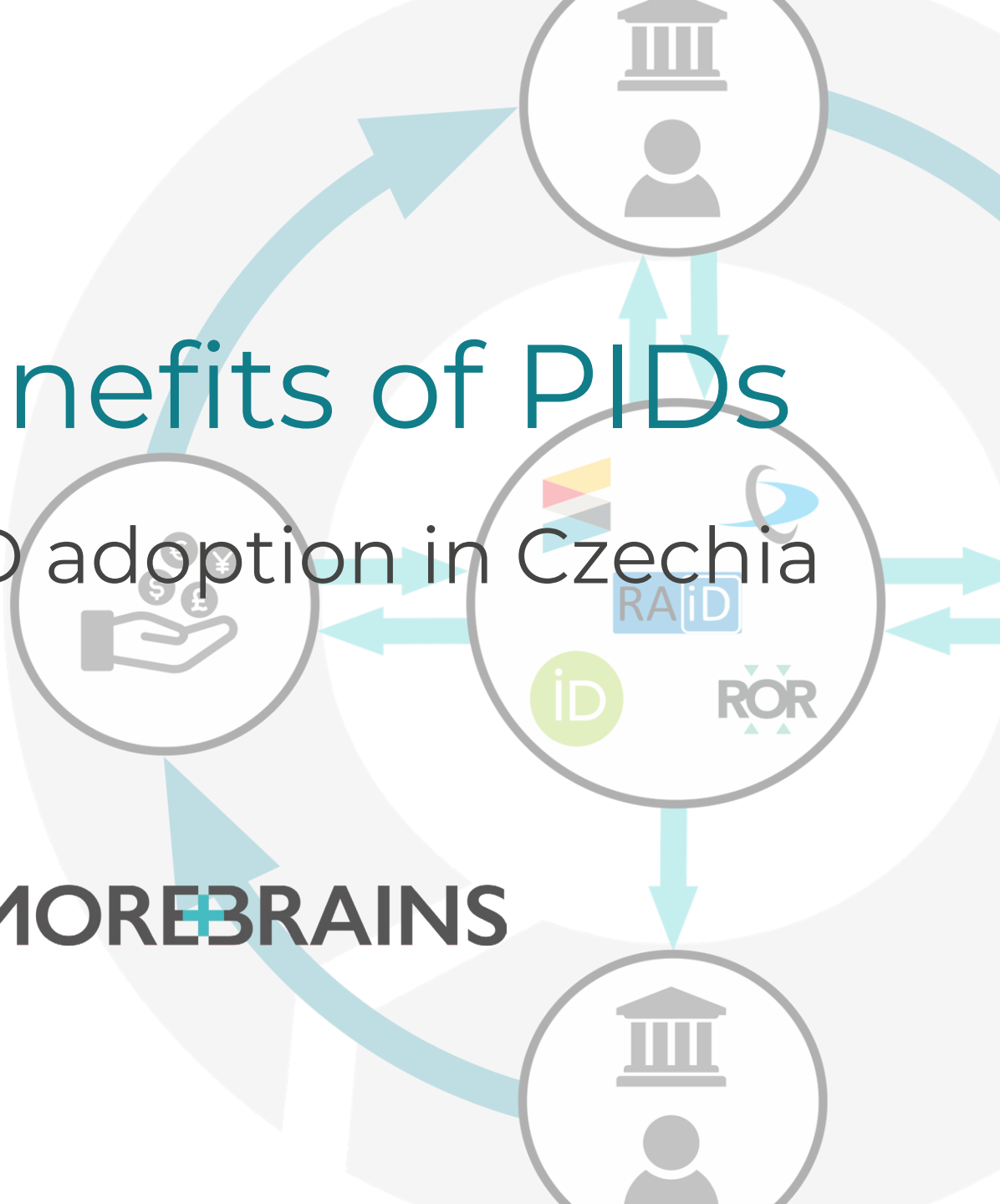
orcid@techlib.cz

doi@techlib.cz

issn@techlib.cz

Measuring the benefits of PIDs

Cost-benefit analysis of PID adoption in Czechia

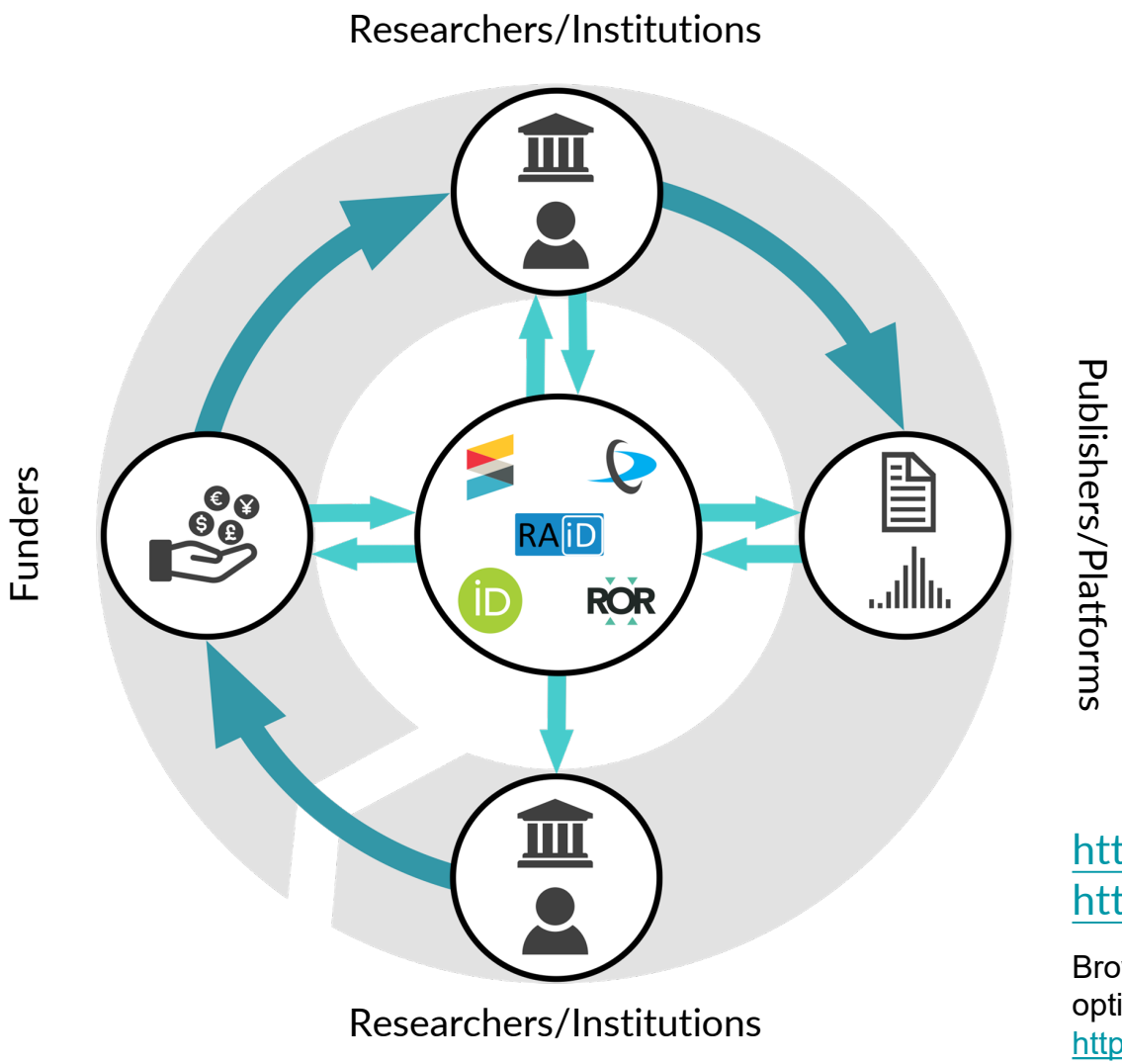


NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

MOREBRAINS

Introduction:
How PIDs save money

PID Optimised research cycle

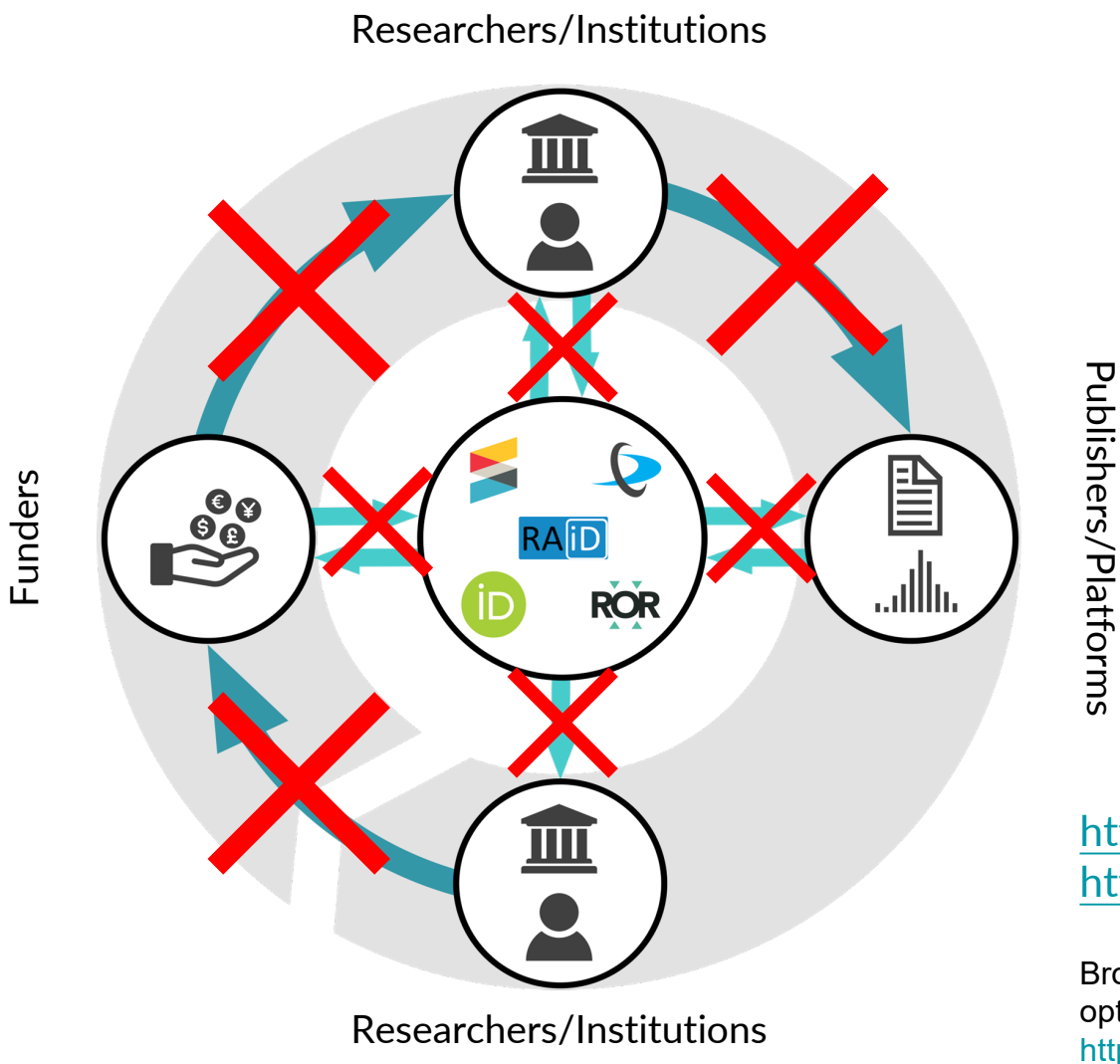


- Open: Information in the form of metadata can move from one system to another
- Efficient: Automated movement of information results in complete and timely records
- Trackable: Links between PIDs enable connections to be analysed
- Requires standards, technical and social integrations between
  - Funders
  - Research Institutions
  - Publishers

https://resources.morebrains.coop/pidcycle/
https://doi.org/10.5281/zenodo.4991733

Brown, Josh, Jones, Phill, Meadows, Alice, & Murphy, Fiona. (2022, September 16). PID-optimised workflows: A vision of a more efficient future. Zenodo. https://doi.org/10.5281/zenodo.7085489

PID Optimised research cycle



- Open: Information in the form of metadata can move from one system to another
- Efficient: Automated movement of information results in complete and timely records
- Trackable: Links between PIDs enable connections to be analysed
- Requires standards, technical and social integrations between
  - Funders
  - Research Institutions
  - Publishers

https://resources.morebrains.coop/pidcycle/
https://doi.org/10.5281/zenodo.4991733

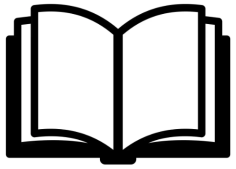
Brown, Josh, Jones, Phill, Meadows, Alice, & Murphy, Fiona. (2022, September 16). PID-optimised workflows: A vision of a more efficient future. Zenodo. https://doi.org/10.5281/zenodo.7085489

Wasted time and effort duplication

Information about Research activity

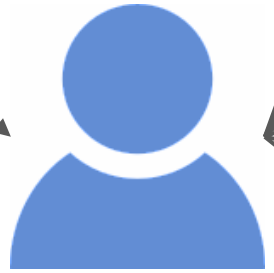
Grant

Data



Publication

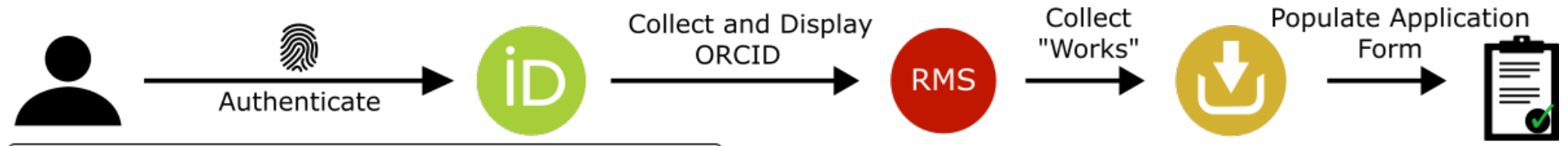
Researcher

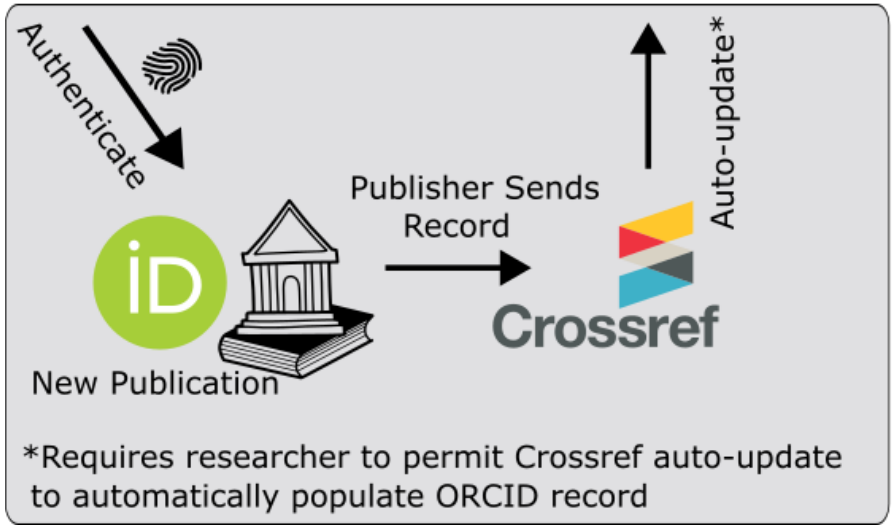


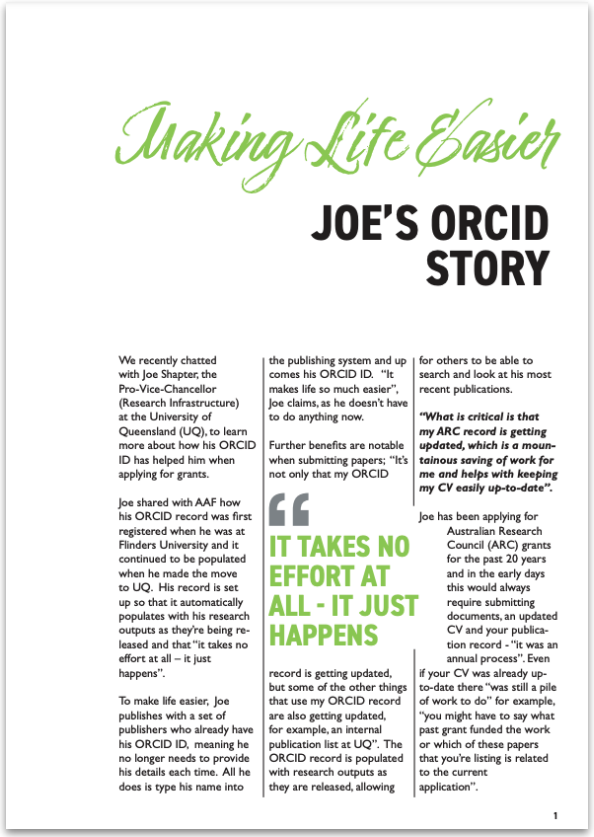
Administrator

- Repositories
- Research information systems
- Funding application systems
- Funder reporting platforms
- Grant management systems
- Finance/HR systems
- Publisher submission systems

Case study: ORCID at Australian Research Council







Making Life Easier

JOE'S ORCID STORY

We recently chatted with Joe Shapter, the Pro-Vice-Chancellor (Research Infrastructure) at the University of Queensland (UQ), to learn more about how his ORCID ID has helped him when applying for grants.

Joe shared with AAF how his ORCID record was first registered when he was at Flinders University and it continued to be populated when he made the move to UQ. His record is set up so that it automatically populates with his research outputs as they're being released and that "it takes no effort at all – it just happens".

To make life easier, Joe publishes with a set of publishers who already have his ORCID ID, meaning he no longer needs to provide his details each time. All he does is type his name into the publishing system and up comes his ORCID ID. "It makes life so much easier", Joe claims, as he doesn't have to do anything now.

Further benefits are notable when submitting papers; "It's not only that my ORCID

"IT TAKES NO EFFORT AT ALL - IT JUST HAPPENS

record is getting updated, but some of the other things that use my ORCID record are also getting updated, for example, an internal publication list at UQ". The ORCID record is populated with research outputs as they are released, allowing for others to be able to search and look at his most recent publications.

"What is critical is that my ARC record is getting updated, which is a mountainous saving of work for me and helps with keeping my CV easily up-to-date".

Joe has been applying for Australian Research Council (ARC) grants for the past 20 years and in the early days this would always require submitting documents, an updated CV and your publication record - "it was an annual process". Even if your CV was already up-to-date there "was still a pile of work to do" for example, "you might have to say what past grant funded the work or which of these papers that you're listing is related to the current application".

1

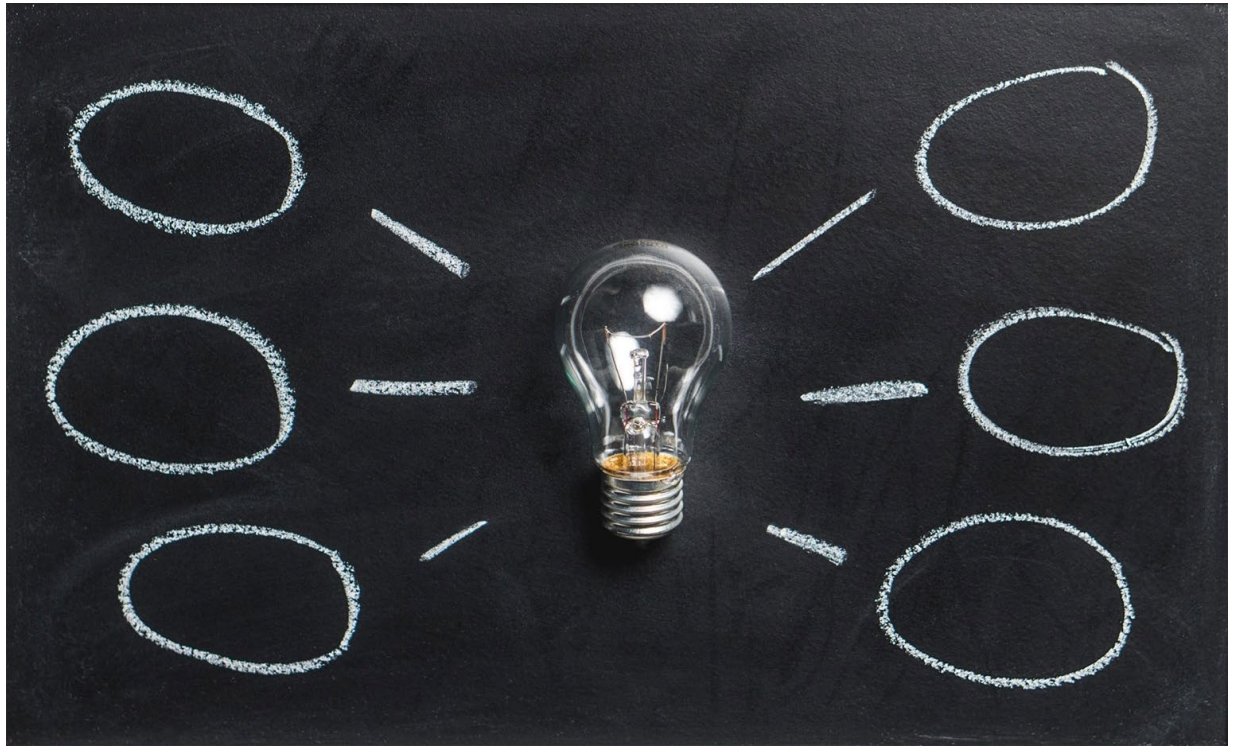
This saved me 3-4 days per grant application - the difference in workload was staggering!"

Joe Shapter, Pro-Vice-Chancellor (Research Infrastructure) University of Queensland

The benefits of PIDs for Czech policy priorities

What are the benefits of PIDs for Czechia?

PIDs are widely recognised as a fundamental component of a modern digital research system.



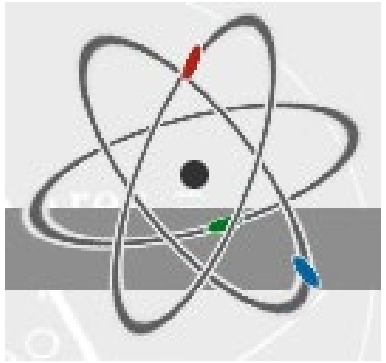
More complete, accurate, and timely information about research activities, outcomes, and impacts are critical to a number of priority areas for the Czech government and people.

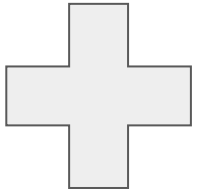
PIDs will serve the National Research and Innovation Strategy for Smart Specialization of the Czech Republic 2021 - 2027 (the RIS3 strategy) and the National Research, Development and Innovation Policy of the Czech Republic 2021+ for example.

PIDs create efficiencies

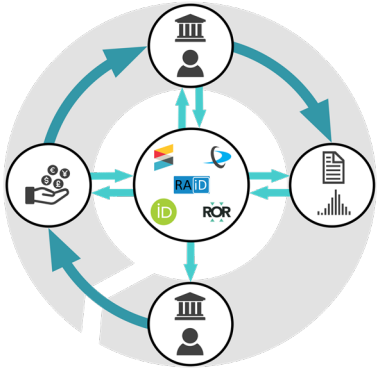
The national IS VaVaI platform is the primary destination for research reporting in the Czech Republic. It delivers efficiencies via:

- Reuse of data already entered once in other public administration information systems
- Cooperation in sharing data initially entered into IS VaVaI





The re-use of metadata from PIDs is a valuable complement to this proven approach, ***enabling improvements to efficiency within institutions***.



These efficiencies reinforce the value of existing investments, such as in the continuing development of IS VaVaI, as well as serving to justify investment in the integration of PID APIs into Czech systems.

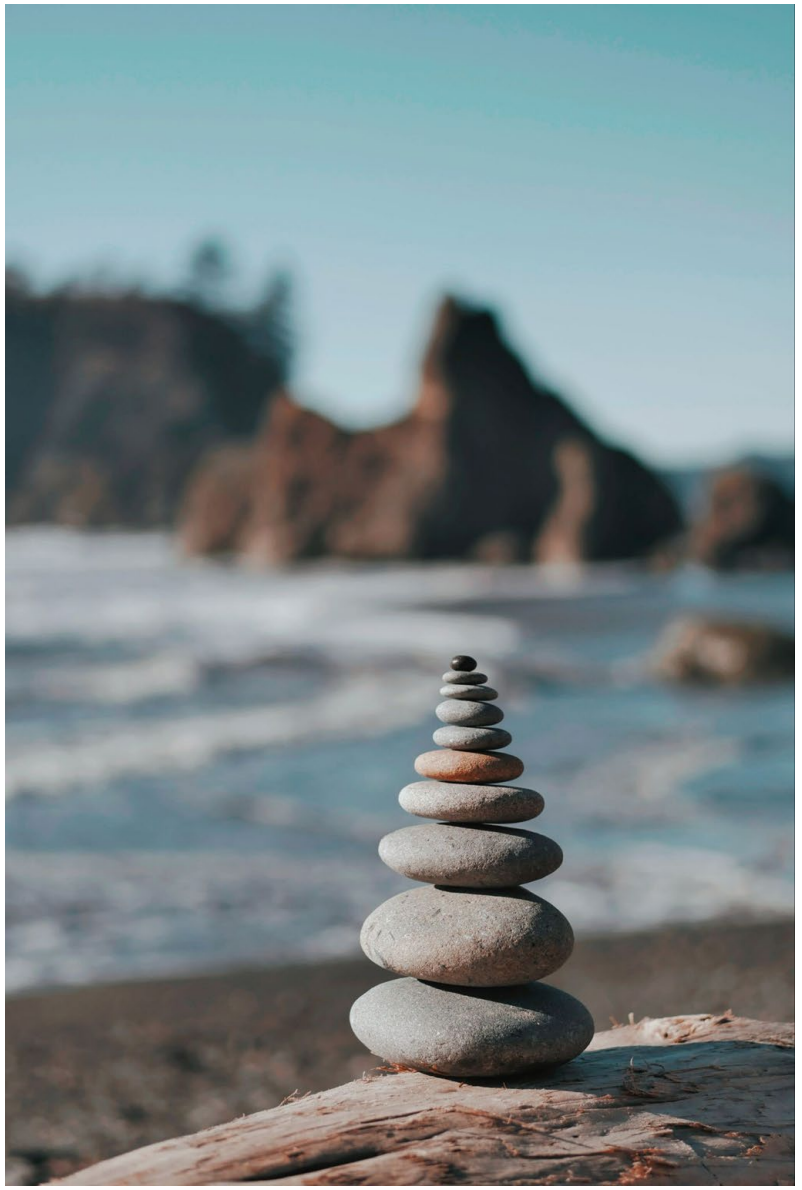
PIDs support open research

PIDs are a fundamental building block for open research.

They underpin plan S and the European Open Science Cloud.

European Commission guidance on access to scientific research states: *"Member States should ensure that… publications resulting from public funding are easily identifiable by appropriate technical means, including through metadata attached to electronic versions of the research output and persistent identifiers."*

https://eur-lex.europa.eu/legal-content/CS/TXT/PDF/?uri=CELEX:32018H0790&from=EN



Data and methodology

The mechanism for direct cost savings

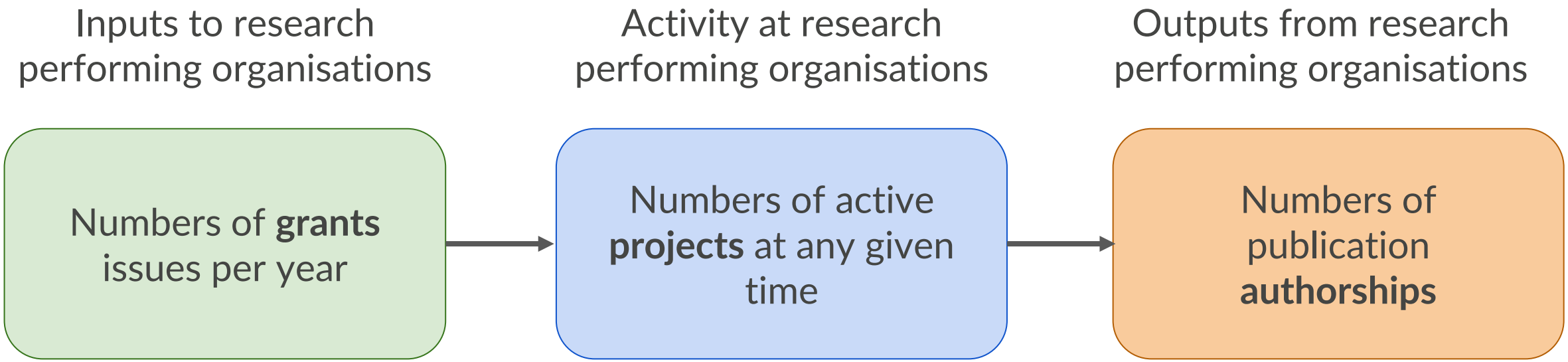
Researchers and administrators spend significant amounts of time keying metadata into various systems:

- Repositories
- Research information systems
- Funding application systems
- Funder reporting platforms
- Grant management systems
- Finance / HR systems
- Publisher submission systems
- And more…

Question: How much time (and therefore money) would be saved if this information about research activity moved from system to system automatically?

Estimating the scale of research activity

We looked at three major quantifiable entities of research activity

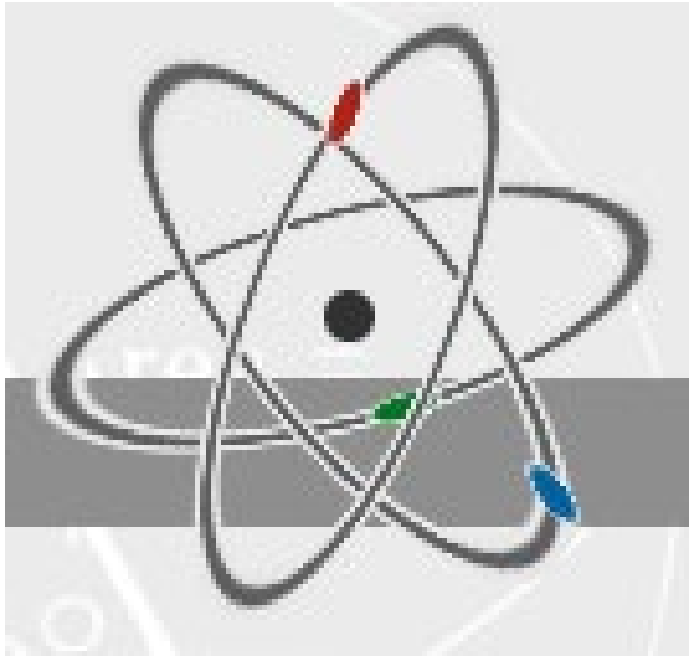


The primary data sources (Grants and Authorships)

The Czech national research information database (IS VaVaI) gathers, processes, and discloses data on R&D activities that are Czech public-funded.

IS VaVal consists of four modules:

- Database of programmes (CEA)
- Database of calls for proposals (VES)
- **Database of awarded grants (CEP)**
- **Database of research outputs (RIV)**



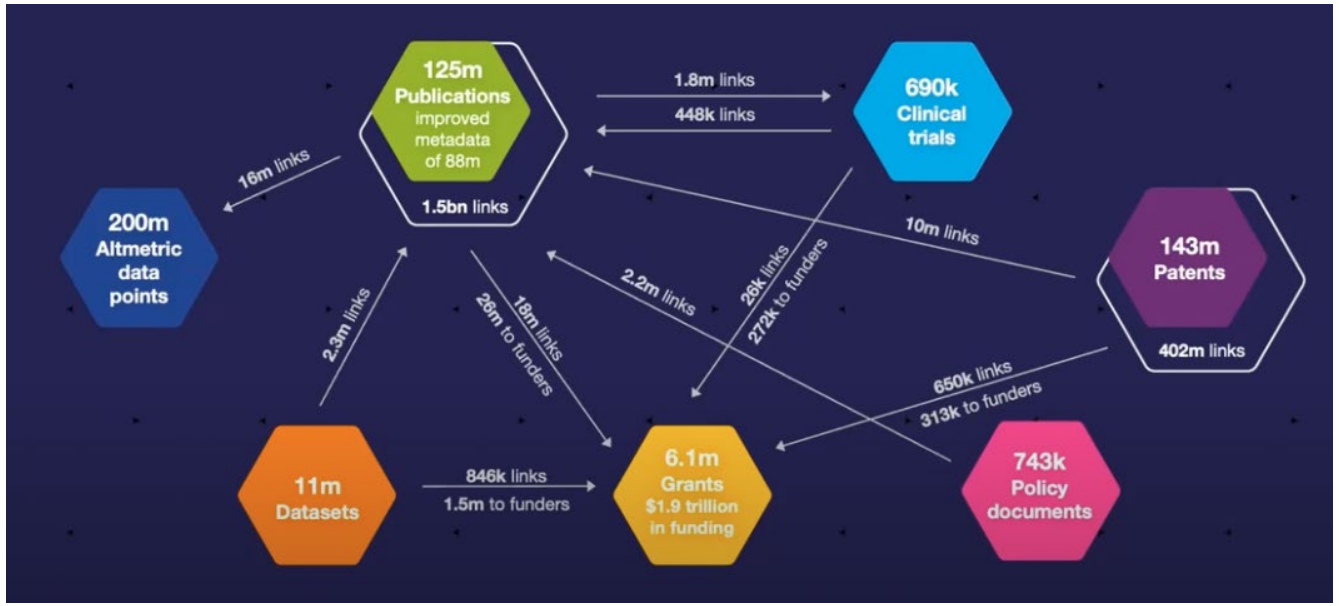
Supplemental data from Digital Science Dimensions

EU funding - contains data that is in CORDIS

- Horizon Europe
- Horizon 2020
- Etc.

Contains data from non-European funders

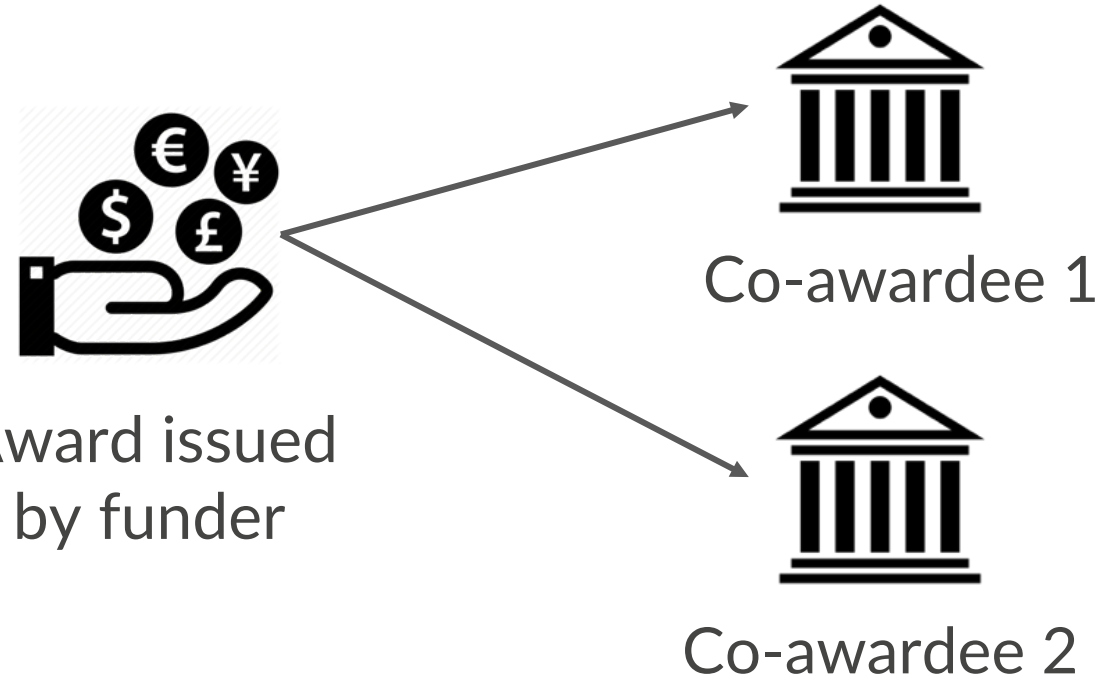
We also used Dimensions to identify Czech publications not funded by Czech funders



Number of grant awardees

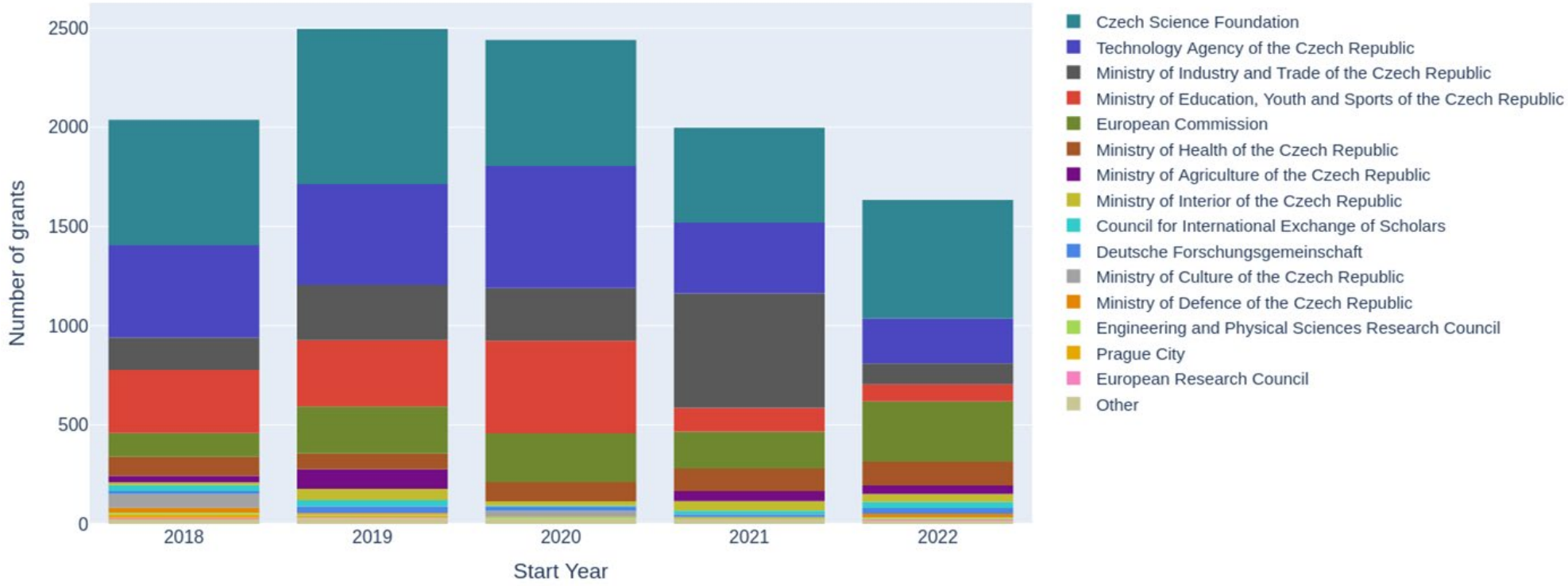
We need to know how many grants were awarded to each institution in Czechia

For example, a grant that is awarded to two Czech institutions is counted twice



- **Database of awarded grants (CEP)**
- **Supplemented by Dimensions**
- Each institution must manage its own grant information
- We therefore counted once for each institution supported by each grant

2,000 funding awards



- Primary source of awarded grants data was the CEP database
- Augmented by Dimensions data to give a more complete picture including Horizon Europe and other funding included in the CORDIS database

Estimating the number of projects

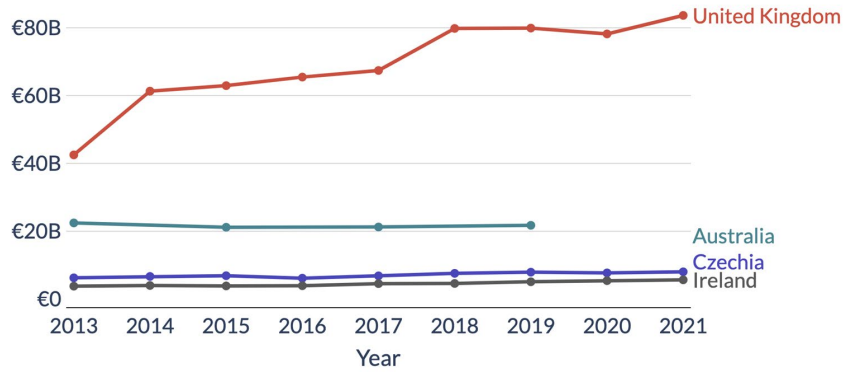
Dr Simon Kerridge, DProf
Director of Research Services, University of Kent
Chair, Association of Research Managers and Administrators (UK)

- No global or national databases exist, so we have to estimate the number
- Assumption: Number of projects is proportional to the size of the research environment

How do we estimate?

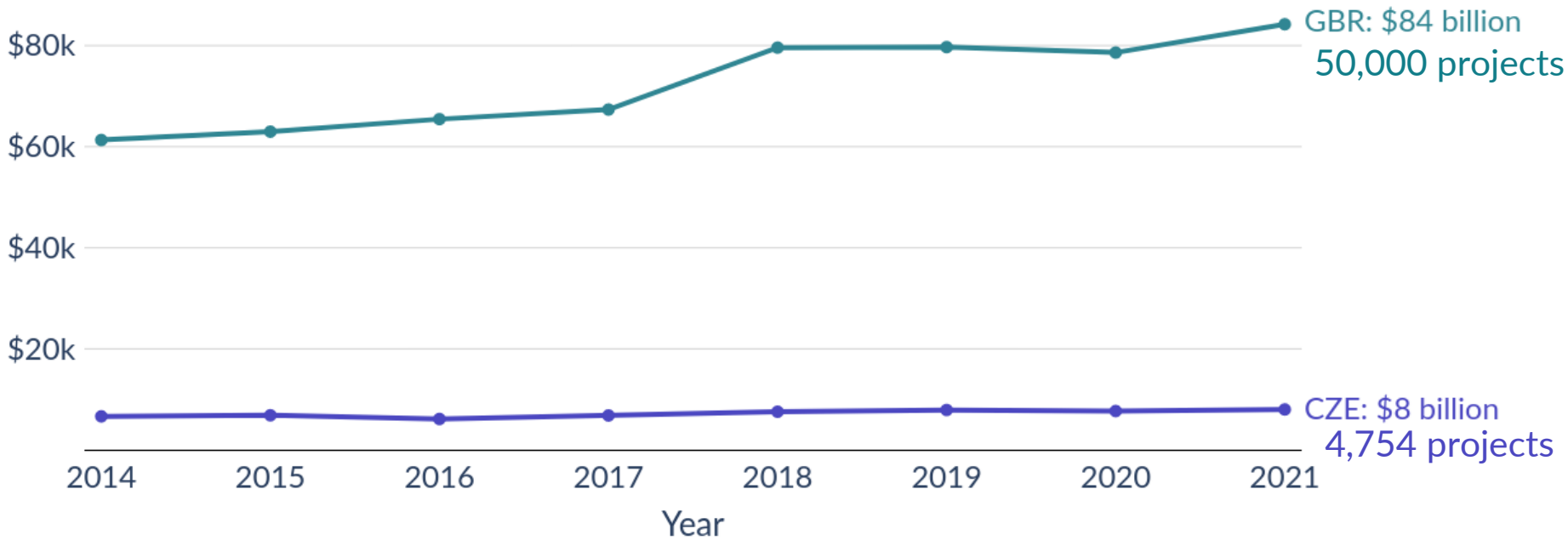
- Benchmarked against estimated number of projects in the UK from Simon Kerridge
- Scaled by total R&D spend according to OECD*

$$(\text{N projects in UK}) \times \frac{(\text{R\&D spend in country of interest})}{(\text{R\&D spend in UK})}$$



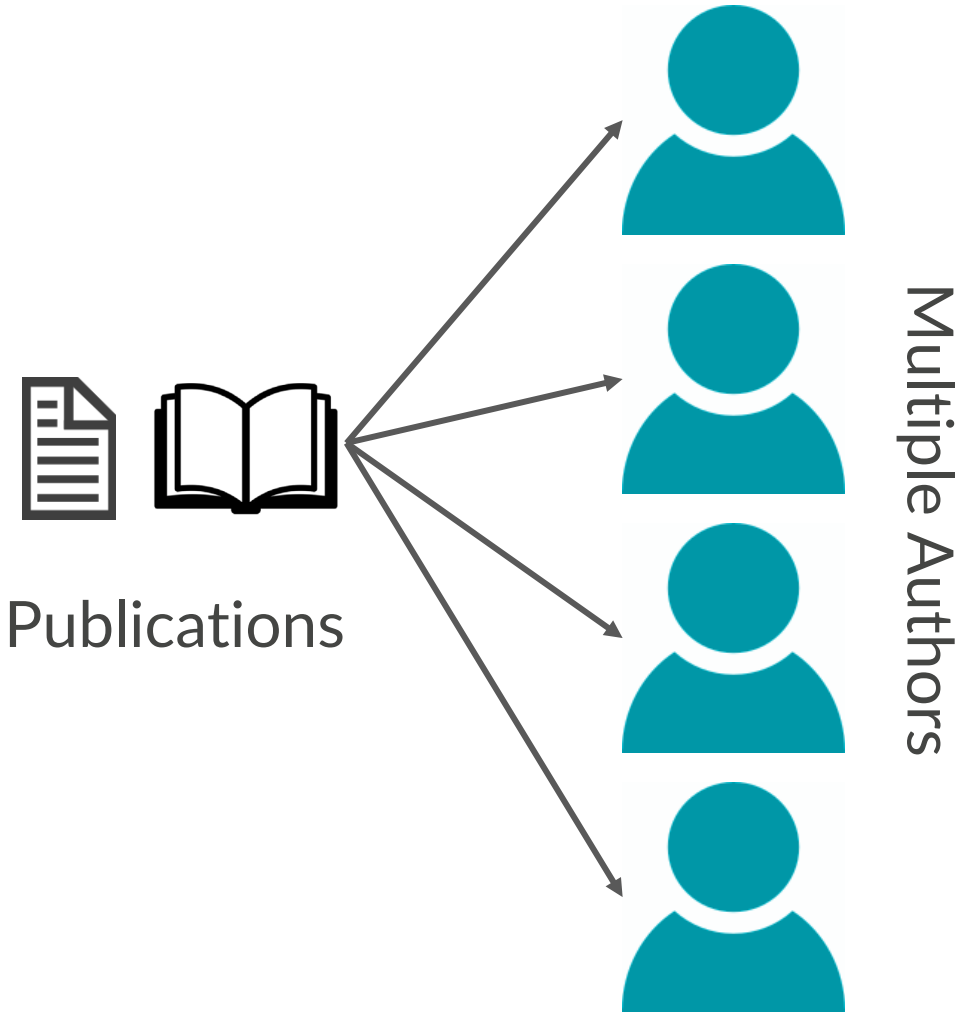
* https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm

Nearly five thousand active projects per year



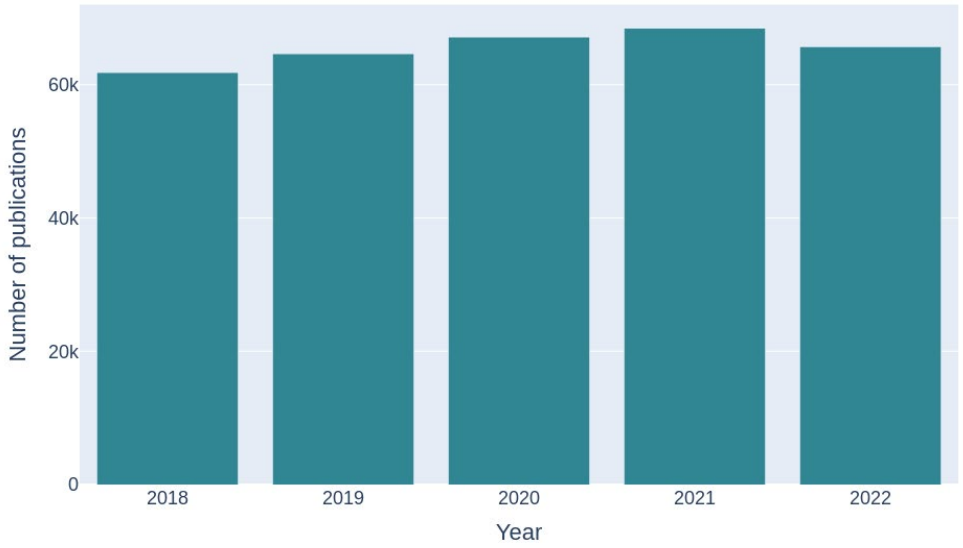
- Benchmark against estimated number of projects in the UK (as per previous CBAs)
- Scaling was based on levels of research funding from OECD statistics

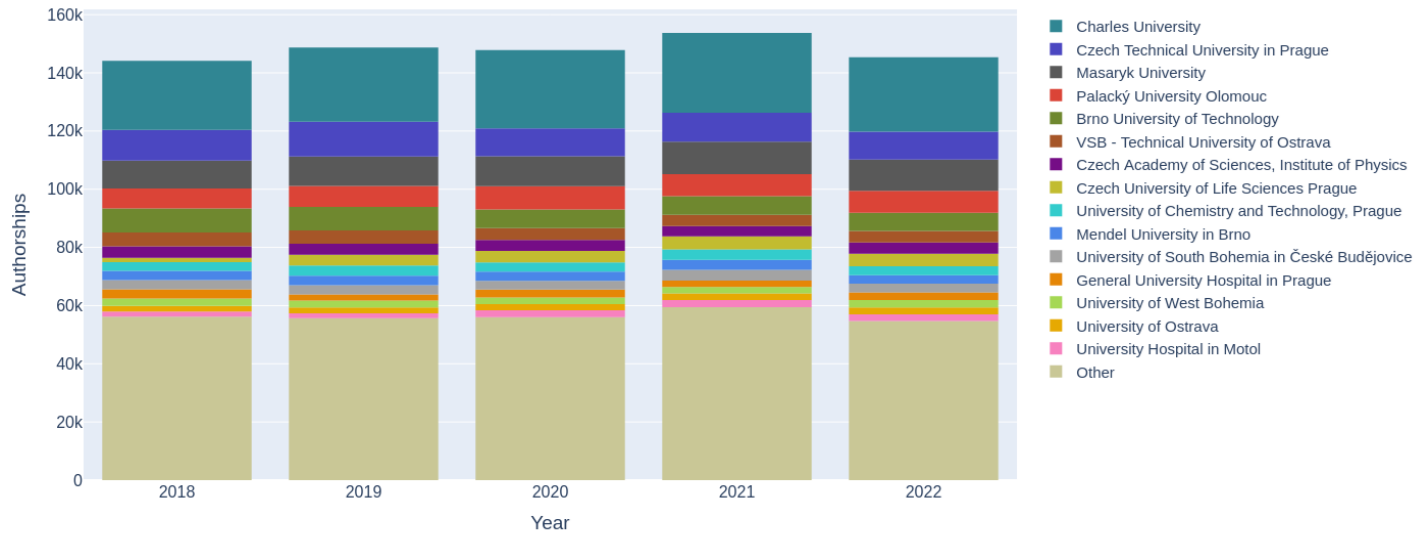
Publication data



- **Database of research outputs (RIV)**
- **Supplemented by Dimensions**
- Each author has reporting responsibilities, resulting in duplication of effort
- We therefore counted once for each author who is affiliated with a Czech institution

67,000 publications per year; 146,000 authorships





- Primary source of awarded grants was the RIV database
- Supplemented with Dimensions data to find more publications from Czech institutions that are funded by non Czech funders

How much time does all this take?

How many rekey events take place?

Number of times publication and grant metadata is rekeyed on average:

- 3.1 and 3.25 times respectively, based on a previous MoreBrains survey *1

Number of times project data is rekeyed:

- 6 times, based on previous third-party research *2

How long does each event take?

Time to input publication metadata:

- 6m 43s, based on previous third-party research *3

Time to input grant and project metadata:

- 10m, based on previous third-party research *2

*1 J. Brown, P. Jones, A. Meadows and F. Murphy, 'Incentives to invest in identifiers: A cost-benefit analysis of persistent identifiers in Australian research systems'. Zenodo, Sep. 30, 2022. doi: 10.5281/zenodo.7100578
*2 Klausen (2017) http://dx.doi.org/10.1016/j.procs.2017.03.011
*3 R. Johnson, H. Henderson, and H. Woodward, 'Institutional ORCID Implementation and Cost- Benefit Analysis Report', Jisc-ARMA, Jul. 2015. [Online]. Available: https://doi.org/10.5281/zenodo.1445290

Source for the salary data

- Data from Czech ministry of labour and social affairs national survey on average earning

Who enters metadata into systems and how much does their time cost?

We assumed that the work is done by the following positions

- Teachers at universities and colleges (ISCO 2310).
- Professionals in administration and organisational administration (ISCO 3343)

https://www.ispv.cz/cz/Vysledky-setreni/Aktualni.aspx

Time burden is multiplied by salaries / overheads

$\sum$ *Over each entity type* (N of entities) × (N of people or institutions) × (time to enter in multiple systems) × (salary & overhead)

1 - Calculate the number of entities
2 - Multiply by the number of people / institutions
3 - Multiply by the number of systems and time taken
4 - Multiply by salary data

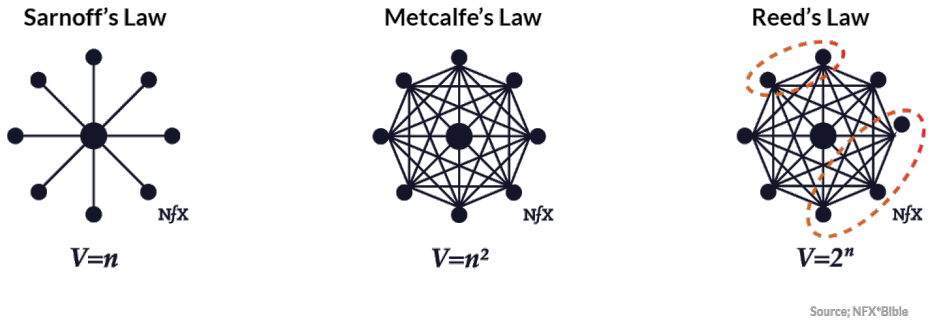
Findings

Total potential savings available

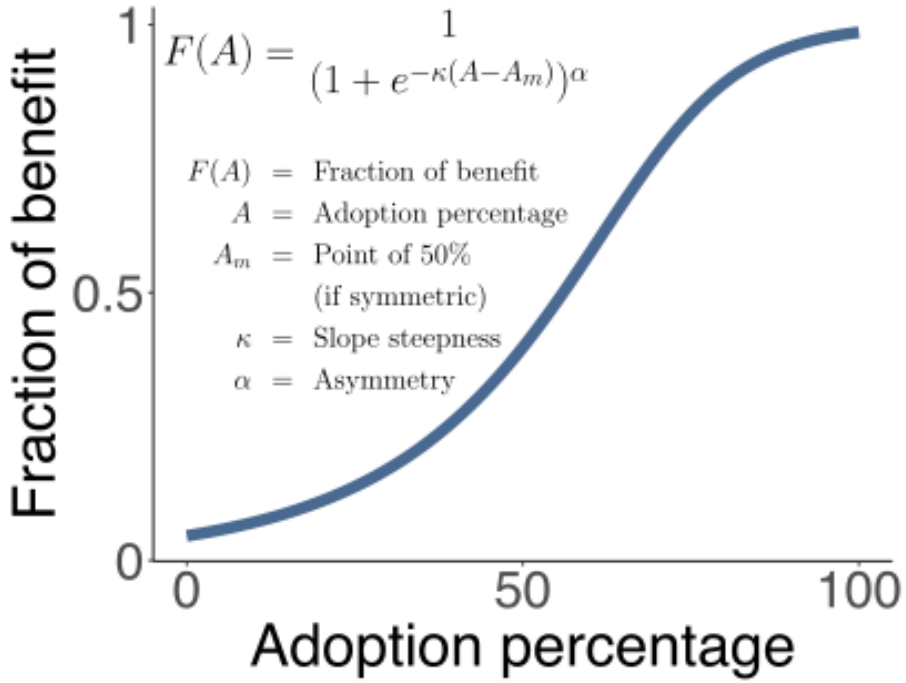
Potential total sector savings							
	Number	*# rekey events*	*# minutes / event*	*Time savings per year (person days)*	*cost / author / minute*		*Financial savings per year*
Publication metadata	148,992	3.1	6.73	**6,476**	8.58 CZK	- if research administrator	26,666,605 CZK
					10.45 CZK	- if researcher	32,482,306 CZK
						Average	**29,574,456 CZK**
Grant metadata	2,024	3.25	10	**137**	8.58 CZK	- if research administrator	564,315 CZK
					10.45 CZK	- if researcher	687,387 CZK
						Average	**625,851 CZK**
Project descriptions	4,754	6	10	**594**	8.58 CZK	- if research administrator	2,447,128 CZK
					10.45 CZK	- if researcher	2,980,821 CZK
						Average	**2,713,975 CZK**
Total predicted annual savings from auto-feed of key metadata via API links:				**7,207**			**32,914,281 CZK**

- Savings that would be created if ALL data re-entry was automated
- Gross savings, not offset against costs of implementation

Benefits depend on the proportion of adoption



- PIDs are network entities - network effects
- The value of adopting a PID increases when other organisations do so
- We use logistic (lazy S) function to model the relationship between adoption level and value*



* D. Kucharavy and R. De Guio, 'Application of S-shaped curves', Procedia Engineering, vol. 9, pp. 559–572, 2011, doi: 10.1016/j.proeng.2011.03.142.

Net savings based on levels of adoption

Net savings based on levels of adoption								
Adoption levels	**0%**	**20%**	**40%**	**50%**	**55%**	**60%**	**80%**	**90%**
Effective benefit (considering adoption curve)	0.0%	4.7%	26.9%	50.0%	62.2%	73.1%	95.3%	98.2%
Total sector time savings in person-days every year	0	342	1938	3604	4486	5269	6865	7078
Cost savings (benefit) every year	-	1,560,989 CZK	8,852,014 CZK	16,457,141 CZK	20,487,802 CZK	24,062,268 CZK	31,353,293 CZK	32,322,278 CZK
Scenario 1: With no central support from NTK PID programme								
Cost of Implementation /year across all institutions	(23,483,156 CZK)	(23,483,156 CZK)	(23,483,156 CZK)	(23,483,156 CZK)	(23,483,156 CZK)	(23,483,156 CZK)	(23,483,156 CZK)	(23,483,156 CZK)
Net savings offset by costs	**(23,483,156 CZK)**	**(21,922,167 CZK)**	**(14,631,142 CZK)**	**(7,026,015 CZK)**	**(2,995,354 CZK)**	**579,112 CZK**	**7,870,137 CZK**	**8,839,122 CZK**
Scenario 2: With central support from NTK PID programme								
Cost of Implementation /year across all institutions	(11,741,578 CZK)	(11,741,578 CZK)	(11,741,578 CZK)	(11,741,578 CZK)	(11,741,578 CZK)	(11,741,578 CZK)	(11,741,578 CZK)	(11,741,578 CZK)
Cost of NTK PID Programme	(7,365,987 CZK)	(7,365,987 CZK)	(7,365,987 CZK)	(7,365,987 CZK)	(7,365,987 CZK)	(7,365,987 CZK)	(7,365,987 CZK)	(7,365,987 CZK)
Net savings offset by costs	**(19,107,565 CZK)**	**(17,546,576 CZK)**	**(10,255,551 CZK)**	**(2,650,424 CZK)**	**1,380,237 CZK**	**4,954,703 CZK**	**12,245,728 CZK**	**13,214,713 CZK**

- These calculations can be used to set targets
- Shows financial breakeven at between 50-55% adoption
- Targets in the 60-80% range would yield significant savings
- Support from CARDS will increase levels of adoption of PIDs so targets will be reached faster

Savings created by the consortia

Net savings due to consortia membership			
PID organisation	Number of institutions	Cost per institution	Cost for all institutions
Costs of individual memberships			
DataCite	11	(50,000 CZK)	(550,000 CZK)
ORCID	24	(207,000 CZK)	(4,968,000 CZK)
Total costs			(5,518,000 CZK)
Costs of consortium membership			
DataCite	N/A	N/A	(50,000 CZK)
ORCID	24	(100,740 CZK)	(2,417,760 CZK)
Total costs			(2,467,760 CZK)
Savings per year created by both consortia			**3,050,240 CZK**

- These figures show current sector savings from the two consortia
- Savings will increase if membership increases

Recommendations to drive progress

Understanding PID adoption

PID adoption refers to the use of PIDs and metadata in systems to save time and effort by reusing metadata and automating processes. Adoption relies on two interdependent factors: PID coverage and PID integrations.

PID coverage is vital to encourage PID integrations. If only 5% of researchers have an ORCID ID, there is very little incentive to use resources to integrate the ORCID API into a system.

PID integrations enable the benefits of PIDs to be delivered, allowing metadata to be reused and, therefore, saving researchers and administrators time and money.

The greater the level of PID coverage, the higher the potential value of a PID integration.

Two ways to measure PID coverage

1. Where the total number of instances of a given entity can be established (e.g. the number of publications or grants), we can compare that total with the number of PIDs in the relevant registry or registries. This measure of PID coverage can then be expressed as a **percentage of the total number of known entities**

2. If this information is not available (e.g. we do not have reliable data on the total number of datasets at a given time), we can, instead, track changes in the number of PIDs registered for each entity over time, by measuring **the absolute number registered and the rates of registration per year for each**

This information can then be used to help to set realistic targets for improvements in coverage.

How to measure PID integration

A PID integration is when a software instance has an active connection to a PID registry API and is reading metadata from the registry, adding metadata to records, or both.

Integration levels are measured by counting **the number of active API connections in use at the time of investigation**.

We do not currently have access to data to evaluate the levels of PID integrations.

The only reliable method that has been developed for assessing integrations is to use surveys.

Recommendations

Our recommended next steps are:

1. Undertake follow-up research – to survey levels of current and planned PID integrations in Czech institutions and to evaluate the coverage of Czech research entities within PID registries
2. Set a medium to longer term target of 80% coverage across all five priority PIDs, reinforced with a target of 60% institutional integration of PIDs
3. Promote the value of PIDs to Czech institutions and researchers to maximise adoption and coverage of key priority PIDs
4. Increase the visibility of Czech research in global PID registries
5. Implement PIDs in shared national systems
6. Undertake additional research to identify keystone systems in which research information is aggregated

Thank you!

Any Questions?